

Figure 1

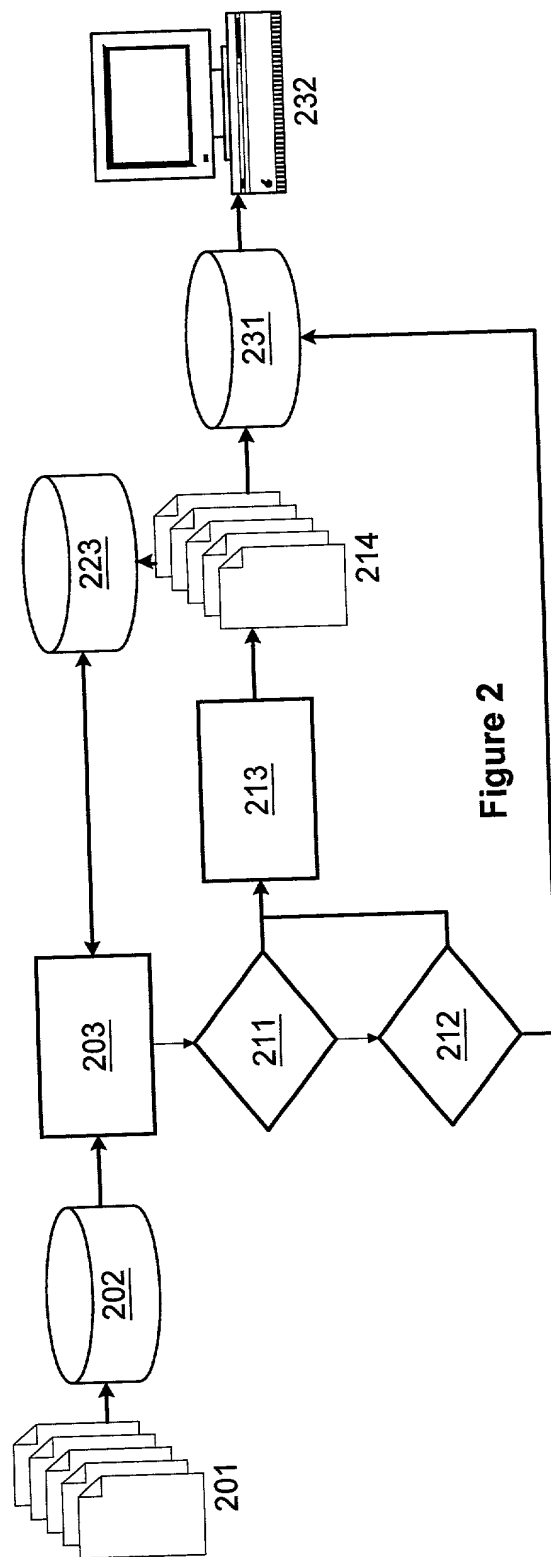


Figure 2

Sheet 2/5

```

<!DOC TYPE registerDoc [
  <!ELE MENT collection ( textitem+)> 301
  <!ELE MENT textitem ((file|text), categories+)> 302
  <!ATT LIST textitem
    extid ID #REQUIRED
    date CDATA #REQUIRED >
  <!ELE MENT text (#PCDATA A)> 303
  <!ELE MENT file EMPTY> 304
  <!ATT LIST file
    location CDATA #REQUIRED
  <!ELE MENT categories ( codeclass+)> 305
  <!ELE MENT codeclass (code+) > 306
  <!ATT LIST codeset
    extid CDATA #REQUIRED >
  <!ELE MENT code EMPTY> 307
  <!ATT LIST code
    extid CDATA #REQUIRED
    date CDATA #REQUIRED >
]

```

Figure 3A

```

<?xml version="1.0" encoding="ISO8859-1"?> 301
<collection> 302
  <textitem extid="ddd" date="2001-05-29"> 303
    <text> 303
      This is the text that Categorizer will use when identifying
      examples to assist it in classification
      of new documents.
    </text>
    <categories> 305
      <codeclass extid="main"> 306
        <code extid="X1" date="2001-05-29T15:34z"/> 307
      </codeclass>
    </categories>
  </textitem>
</collection>

```

Figure 3B

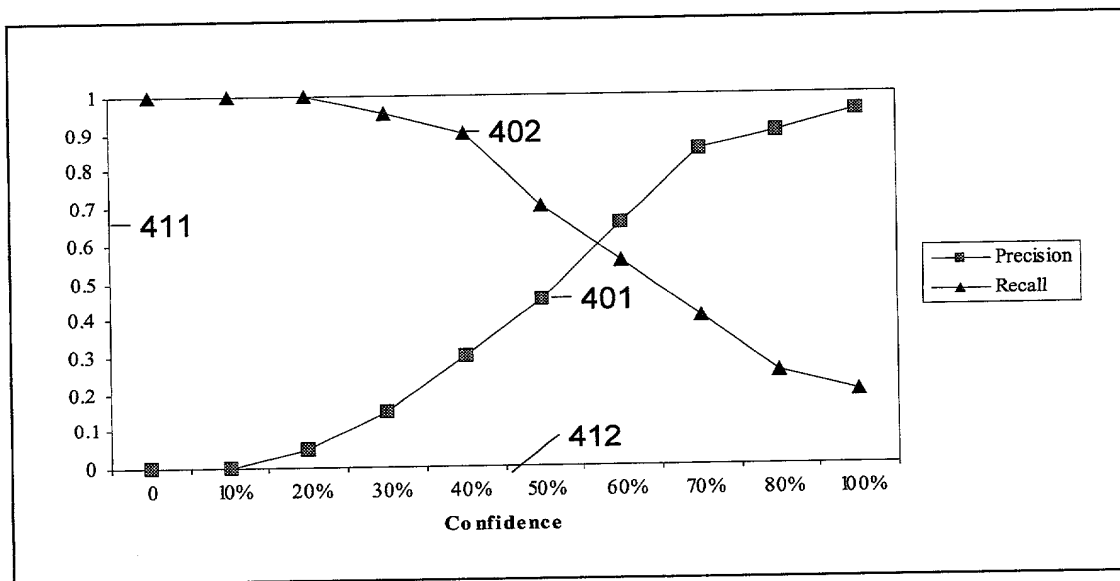


Figure 4

Sheet 2/5

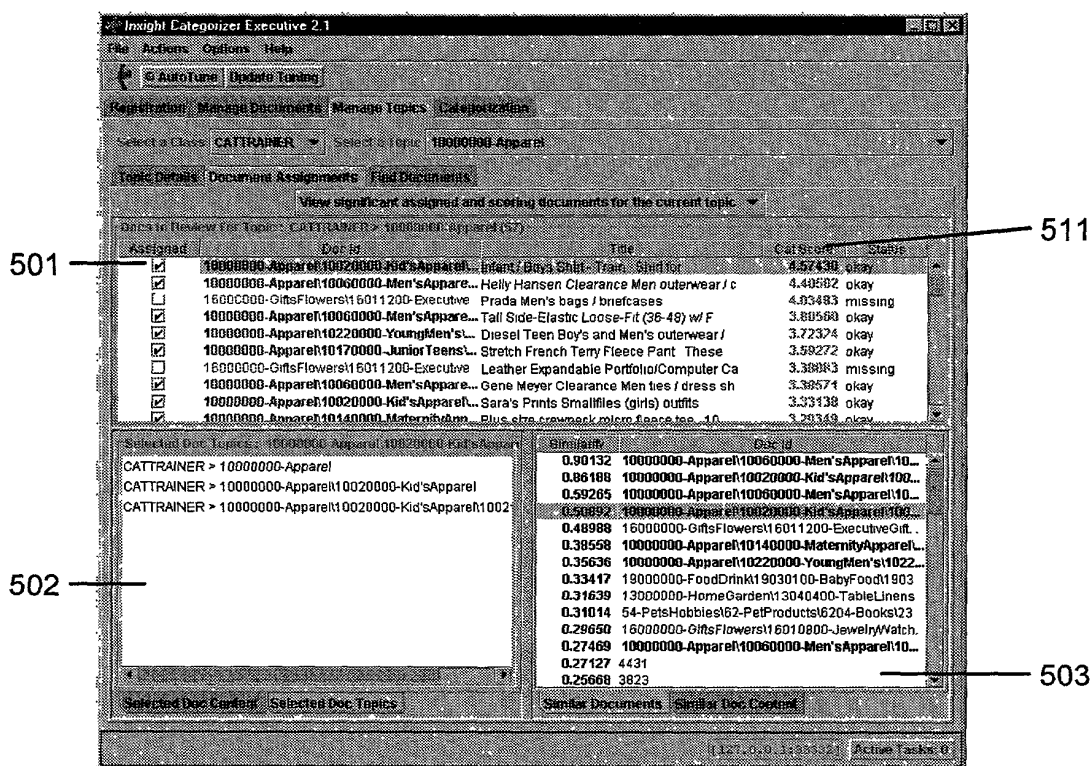


Figure 5A

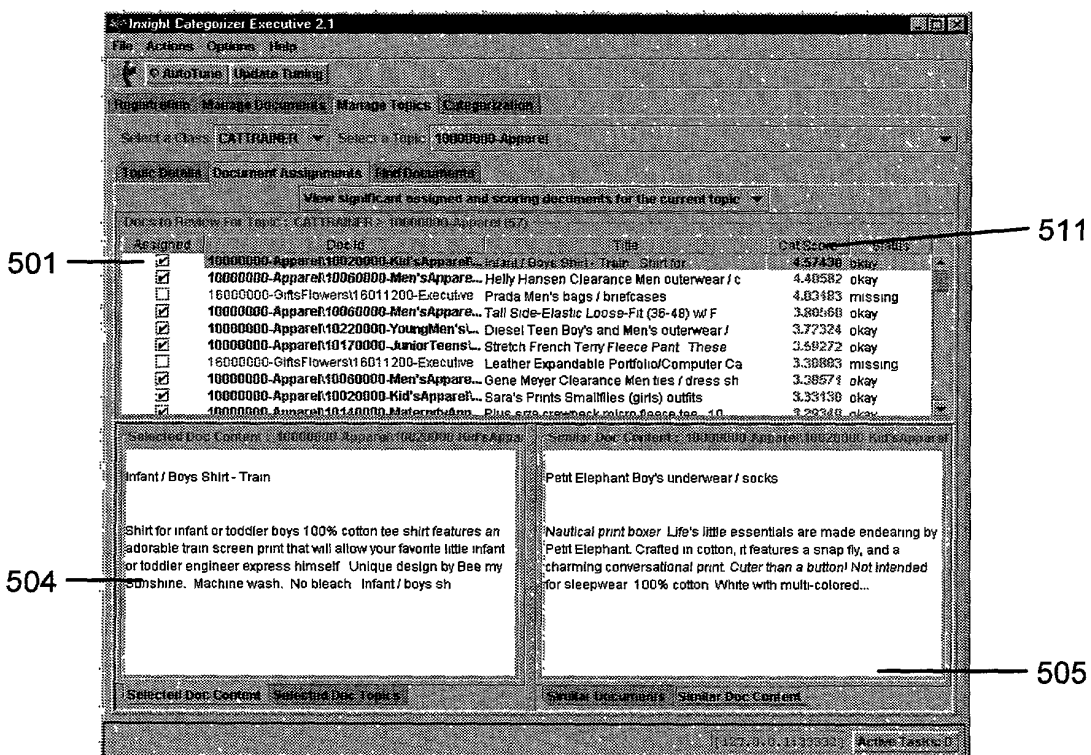


Figure 5B

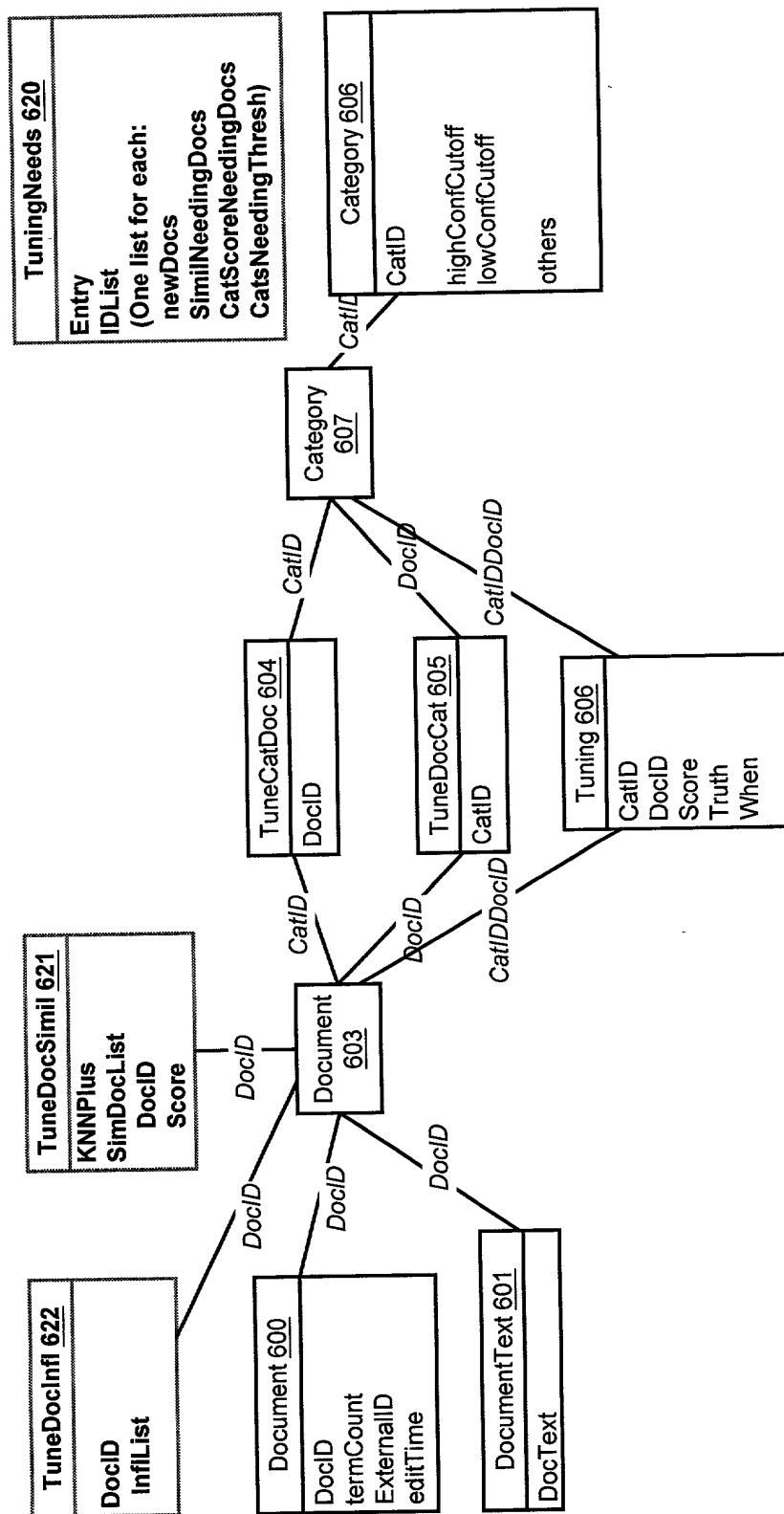


Figure 6

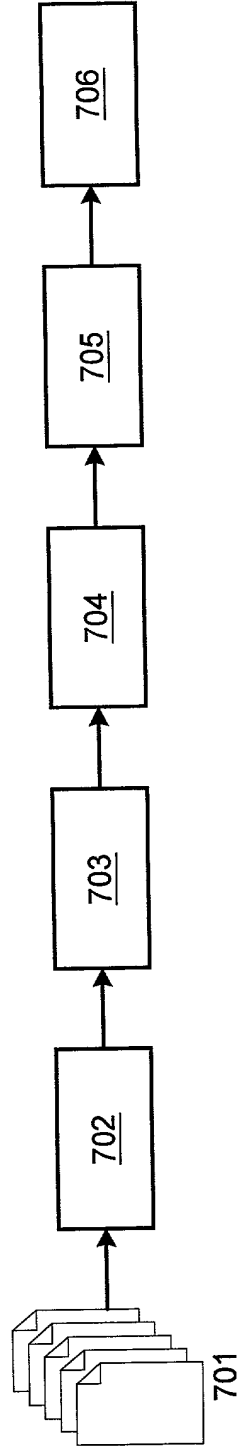


Figure 7